Eur Biophys J (2012) 41:1–11
DOI 10.1007/s00249-011-0751-y

ORIGINAL PAPER

EBSA
Biophysics in Europe

A simple procedure to evaluate the efficiency of bio-macromolecular rigid-body refinement by small-angle scattering

Frank Gabel

Received: 23 June 2011 / Accepted: 6 September 2011 / Published online: 24 September 2011

Abstract A simple and rapid procedure is presented that enables evaluation and visualization of refinement efficiency for bio-macromolecular complexes consisting of two subunits in a given orientation by using small-angle scattering. Subunit orientations within a complex can be provided in practice by NMR residual dipolar couplings, an approach that has been combined with increasing success to complement small-angle data. The procedure is illustrated by applying it to several systems composed of two simple geometric bodies (ellipsoids) and to protein complexes from the protein data bank that vary in subunit size and anisometry. The effects of the experimental small-angle scattering range (Q-range) and data noise level on the refinement efficiency are investigated and discussed. The procedure can be used in two ways: (1) either as a quick preliminary test to probe the refinement capacity expected for a given bio-macromolecular complex *prior to* sophisticated and time-consuming experiments and data analysis, or (2) as an a posteriori check of the stability and accuracy of a refined model and for illustration of the residual degrees of freedom of the subunit positions that are in agreement with both small-angle data and restraints on subunit orientation (as provided, e.g., by NMR).

Keywords Small-angle X-ray scattering · Small-angle neutron scattering · Nuclear magnetic resonance · Residual dipolar couplings · Protein complex · Modeling

Introduction

Small-angle X-ray (SAXS) and neutron (SANS) scattering provide low-resolution structural information about bio-macromolecules and complexes in solution bridging, with electron microscopy, the gap between atomic resolution (crystallography and NMR) and cellular resolution (fluorescence, light microscopy…). In the last decade, SAXS/SANS bio-macromolecular data analysis has undergone tremendous progress (Heller 2010; Jacques and Trewhella 2010; Lipfert and Doniach 2007; Petoukhov and Svergun 2007; Putnam et al. 2007) and several sophisticated approaches have been developed to treat a number of specific problems: ab initio shape determination, rigid-body modeling, addition and analysis of missing loops and flexible parts, and polydisperse solutions. It is generally accepted that supplementing small-angle scattering (SAS) data with complementary structural restraints improves the accuracy of the structural models and reduces the degrees of freedom and ambiguities possible (Bernadó and Blackledge 2010; Madl et al. 2011; Petoukhov and Svergun 2007; Putnam et al. 2007). Such restraints may be provided, amongst others, by cross-linking (Back et al. 2003), mutational analysis (Zuiderweg 2002), NMR chemical shift perturbations (Dominguez et al. 2003), paramagnetic relaxation enhancement (PRE) (Battiste and Wagner 2000), or computational approaches (Pons et al. 2010).

NMR residual dipolar couplings (RDCs) (Bax 2003; Blackledge 2005; Tjandra and Bax 1997) are particularly well-suited to complementing restraints from SAS: RDCs

Electronic supplementary material The online version of this article (doi:10.1007/s00249-011-0751-y) contains supplementary material, which is available to authorized users.

F. Gabel (✉)
Extremophiles and Large Molecular Assemblies Group,
Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS,
UJF, 41 rue Jules Horowitz, 38027 Grenoble, France
e-mail: frank.gabel@ibs.fr

provide restraints on the respective domain (subunit) *orientations* by exploiting and interpreting the anisotopic tumbling of biomacromolecules dissolved in alignment media (stretched gels, dilute liquid crystals, phages, etc.). The potential of combining SAS translational with RDC orientational restraints has been realized recently and is reflected in a growing number of publications (Macheboeuf et al. 2011; Marino et al. 2006; Mattinen et al. 2002; Schwieters et al. 2010; Takayama et al. 2011; Yuzawa et al. 2004). Several systematic ways of combining RDCs with SAS restraints into a common potential function have been proposed recently (Gabel et al. 2006, 2008; Grishaev et al. 2005; Mareuil et al. 2007; Wang et al. 2009a, b).

Although impressive progress in rigid-body modeling using SAS alone or in combination with other techniques (in particular NMR; Madl et al. 2011) has been accomplished, and ever more sophisticated approaches are being applied, several crucial issues remain poorly addressed by most studies to date, in particular the *accuracy, uniqueness*, and *stability* of the refined models. In other words:

1. How stable are the refined models against errors in the SAS data or the angular range used?
2. Is the conformational space exhaustively sampled?
3. Are the refined models the best ones possible or are they trapped in a side-minimum of the target function?
4. What are residual degrees of freedom for the domain (subunit) positions that are compatible with the SAS data? and
5. How do points 1–4 depend on the geometry of the system, i.e. on the relative sizes, positions and anisometries (deviation from spherical shape) of the domains (subunits)?

In this paper, a simple procedure is proposed for evaluation and visualization of the *accuracy, uniqueness* and *stability* of bio-macromolecular models consisting of two rigid bodies (of known individual high-resolution structures) at a given relative orientation that are refined against SAS data. It is based on the fact that the inter-domain (inter-subunit) distance is determined and fixed for a complex of known radius of gyration composed of two rigid bodies according to the *parallel axes theorem* (Gabel et al. 2006; Goldstein 1977). The procedure is applied to several complexes composed of two ellipsoidal bodies of different orientation and anisometry and to different protein complexes varying in subunit shape, distance, and size. In each case, the procedure scores the possible relative positions of the two domains (subunits) against the SAS data from the target structure complex. The results are presented in the conformational space of the domain center positions (polar and azimuthal angles θ and Φ on the sphere determined by the fixed inter-domain distance) as a function of SAS angular range and noise level. They are discussed as a function of domain (subunit) anisometries and respective orientation.

The procedure can be easily and quickly used to evaluate and visualize the expected quality and uniqueness of structural models *prior to* applying sophisticated and time-consuming refinement methods or even before recording actual SAS and NMR data if hypotheses are made on the subunit orientations and inter-subunit distance. Alternatively, it can be applied to a structural model a posteriori to a sophisticated refinement process in order to judge its uniqueness and stability. Although the procedure has been developed for two-body systems, it can in principle be applied to complexes composed of more bodies by using small-angle neutron scattering, contrast variation, and appropriate labeling schemes (or natural contrast) to focus specifically on two partners in a larger complex while minimizing the contribution from the others (Jacrot 1976). The size limit of particles accessible to the approach is mainly dominated by the NMR experiments. However, it has been demonstrated recently that complexes as large as 150 kDa can be studied by NMR residual dipolar couplings if a priori information on the structure of the subunits is available (Schwieters et al. 2010). It can be hoped that this limit will be pushed further with the ongoing methodological development in the NMR community.

Materials and methods

Preparation of bi-ellipsoidal systems and protein–protein complexes from the PDB

Several two-body systems consisting of rotational ellipsoids of different degrees of anisometry and relative position/orientation were studied (Table 1; Fig. 1; Suppl. Table 1; Suppl. Figs 2 and 6). All single ellipsoids were generated by the software DAMMIN (Svergun 1999), version 5.3, in default user mode, imposing in the setup configuration specific axes ratios and lengths in the initial shape and recovering the shapes prior to the annealing process. The output bodies were ensembles of approximately 1,000 dummy atoms filling the ellipsoidal volumes. The single bodies were shifted to their respective positions by hand using the software MOLMOL (Koradi et al. 1996) to generate the bi-ellipsoidal complexes.

In addition four different two-body protein complexes from the protein data bank (PDB) were selected in order to cover a wide range of subunit geometries, relative sizes, and symmetries (Table 1; Fig. 2): gamma-B-crystallin (1AMM) (Kumaraswamy et al. 1996), survivin (1E31) (Chantalat et al. 2000), T-cell antigen receptor (1TCR) (Garcia et al. 1996), and BCL-XL/BAK peptide complex (1BXL) (Sattler et al. 1997). All structures were used as

Table 1 Overview of refinement data of the different models studied

Protein	PDB code	Molecular weight (kDa)	$\lvert B_{max} - B_{min}\rvert/B_{target}$ (%)	$(\theta/\Phi)_{target}$	R_g (Å)
Gamma-B-crystallin	1AMM	42	6.5	0.42/3.12	16.6
Survivin dimer	1E31	33	13.6	1.58/3.74	28.3
T-cell receptor	1TCR	27	23.8	0.92/5.67	24.1
BCL-BC/Bak complex	1BXL	51	4.9	1.61/1.93	17.9

Biellipsoid	Axes ratio first body (Å:Å:Å)	Axes ratio second body (Å:Å:Å)	$\lvert B_{max} - B_{min}\rvert/B_{target}$ (%)	$(\theta/\Phi)_{target}$	R_g (Å)
Side-by-side	50:25:25	50:25:25	33.0	1.57/4.69	37.4
Rod-like	50:25:25	50:25:25	17.1	1.57/3.05	56.2
T-shaped	50:25:25	50:25:25	12.7	1.57/4.71	45.8
L-shaped	50:25:25	50:25:25	11.2	1.57/3.85	36.4

$\lvert B_{max} - B_{min}\rvert/B_{target}$ illustrates the relative variation of the B-values within the θ/Φ-plane for a given system and is a measure for its anisometry

deposited in the PDB apart from the 1AMM dimer where five linker residues (Pro82 through Gly86) were cut away for all further analyses (generation of models and scoring against SAXS data), to avoid contributions from variable linker conformations.

Exhaustive sampling of the conformational space

The conformational space of the subunit positions was sampled using a recently developed procedure (Gabel et al. 2006, 2008). It is based on the fact that in a system consisting of two rigid bodies of known structure, the inter-subunit distance is uniquely determined and fixed by the radius of gyration of the complex they form (*parallel axes theorem*; Goldstein 1977). Historically, it has been used in the context of *label triangulation* (Engelman and Moore 1972; Hoppe 1973; Kratky and Worthmann 1946) for the mapping of the protein positions within the ribosome subunits by determining their distances pair-wise with SANS (Capel et al. 1987; Hoppe et al. 1975).

In our procedure, both bodies in each system are kept at their relative orientation from the initial (target) structure and their centers of mass (centers of scattering density) are allowed to move on the surface of a sphere of fixed diameter corresponding to the inter-subunit distance from the initial (target) complex. 10,000 random shifts of the two subunits on the sphere were carried out by use of Python and C-shell scripts written by the author (available upon request): a given conformation (= relative position of subunits) is described by the polar ($0 \leq \theta \leq \pi$) and azimuthal ($0 \leq \Phi < 2\pi$) angles of the center-of-mass of the first subunit relative to the complex center-of-mass in the Cartesian coordinate frame of the PDB (Figs. 1, 2; Suppl. Fig. 2).

An option to activate *steric restraints* has been introduced in the procedure that generates the ensemble of structures by use of additional Python scripts. Such restraints may comprise a finite linker length or the impossibility of the two domains to interpenetrate. The "finite linker module" discards any structural solutions where the distance between the beginning and the end of the linker exceeds a user-predefined distance in Ångströms. The "interpenetration module" discards all structural models where any atom of domain 1 is closer than a user-predefined distance (in Ångströms) from any atom of domain 2.

Scoring of models against full SAXS data

Using the software CRYSOL (Svergun et al. 1995) in default mode, the theoretical SAXS curves were calculated from the initial complex structures in the Q-range 0–0.3 Å^{-1}. Two different levels of noise were added to the CRYSOL curves using a hand-written Python script according to $I_{noise}(Q) = I_{CRYSOL}(Q) + K\sqrt{I_{CRYSOL}(Q)}$. K was chosen for individual Q-values randomly between 0.5 and −0.5 for low noise-level data and between 5 and −5 for high noise-level data. Several Q-ranges were prepared for each noise level by imposing upper limits of 0.1, 0.2, and 0.3 Å^{-1}, respectively. These scattering curves served as target data against which the 10,000 randomized models (see preceding section) were scored with CRYSOL using 200 data points. The χ^2-values provided by CRYSOL were plotted in the conformational θ/Φ-space (Figs. 1, 2). Illustrative examples of the SAXS curves and χ^2-fits are provided in Suppl. Figs. 1a and 1b.

Evaluation of models using moments of inter-atomic distances

As an attractive, easy, and rapid alternative to the least χ^2-scoring procedure using the full SAXS curve, an approach

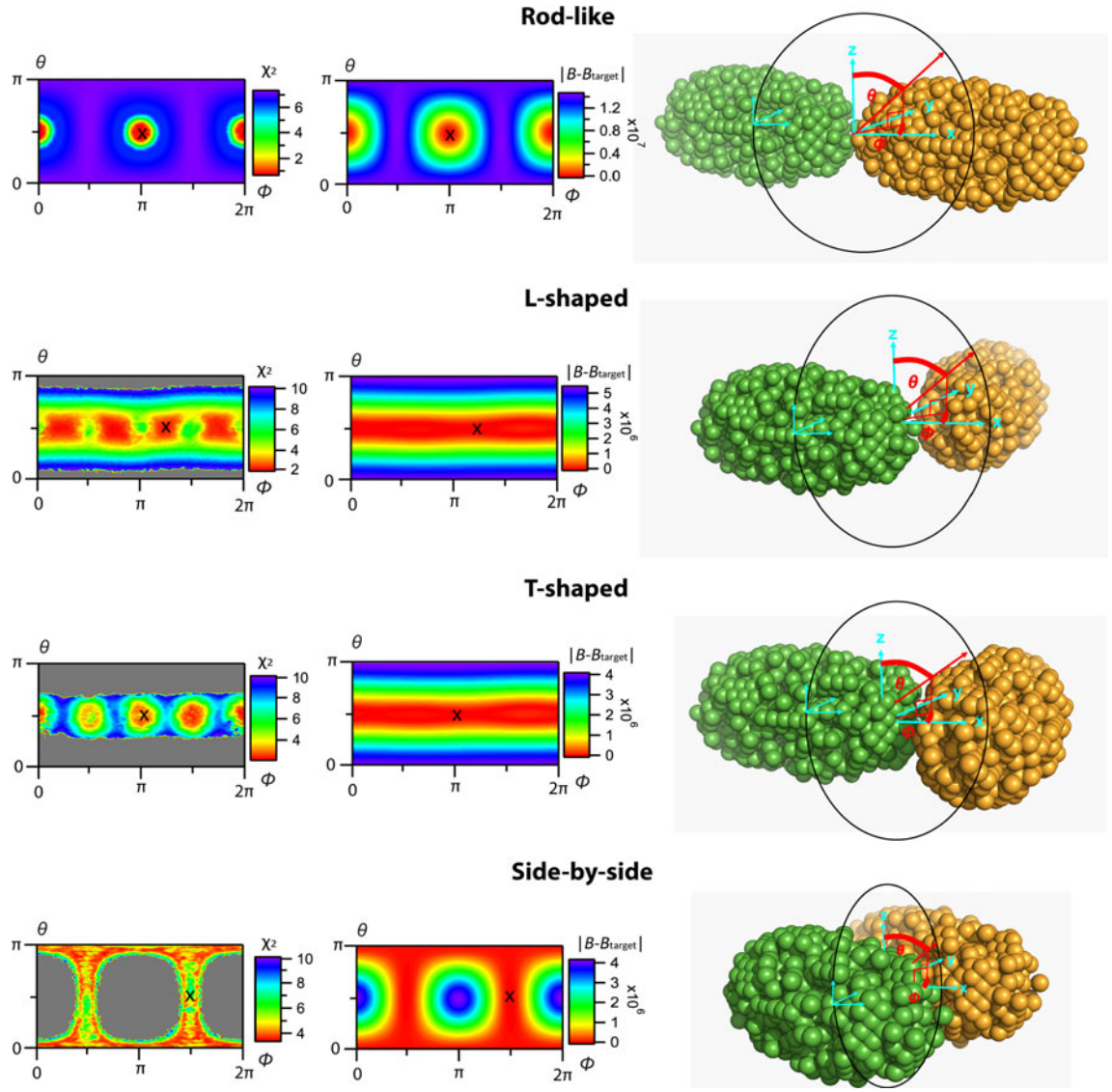

Fig. 1 χ^2 values and $|B - B_{\text{target}}|$ for different bi-ellipsoid systems (Table 1) as a function of the subunit 1 (*green*) center in the θ/Φ-space. *Red regions* indicate lowest χ^2 (*left column*) or $|B - B_{\text{target}}|$ values (*middle column*). The regions in *violet* indicate poorest χ^2 fits or largest $|B - B_{\text{target}}|$ values, respectively. *Grey areas* indicate $\chi^2 > 10$. A *black cross* marks the coordinates of the target structure. The *right column* shows the target structures and the θ/Φ reference frames. The *circles/ellipsoids* indicate, symbolically, the subunit 1 mobility on a spherical surface around the complex center of mass

based on moments of atomic distances within the model is proposed: the SAS curve of a complex consisting of n scattering centers (atoms) at positions r_i can be described in terms of the Debye equation (Debye 1915):

$$I(Q) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j \frac{\sin(Qr_{ij})}{Qr_{ij}} = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j \left[1 - \frac{1}{6}Q^2 r_{ij}^2 + \frac{1}{120}Q^4 r_{ij}^4 - + \cdots\right] \quad (1)$$

with different moments of the atomic coordinates (r_{ij} stands for the distance between the scattering centers i and j and f_i are the scattering lengths of the individual scattering centers):

$$A = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j r_{ij}^2; \quad B = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j r_{ij}^4; \quad C = \sum_{i=1}^{n}\sum_{j=1}^{n} f_i f_j r_{ij}^6. \quad (2)$$

If normalized scattering data are used ($I(Q = 0) = 1$), A can be directly related to the radius of gyration of the complex: $A = 2R_g^2$ (Gabel et al. 2006). As discussed in the same reference the second moment B of the scattering

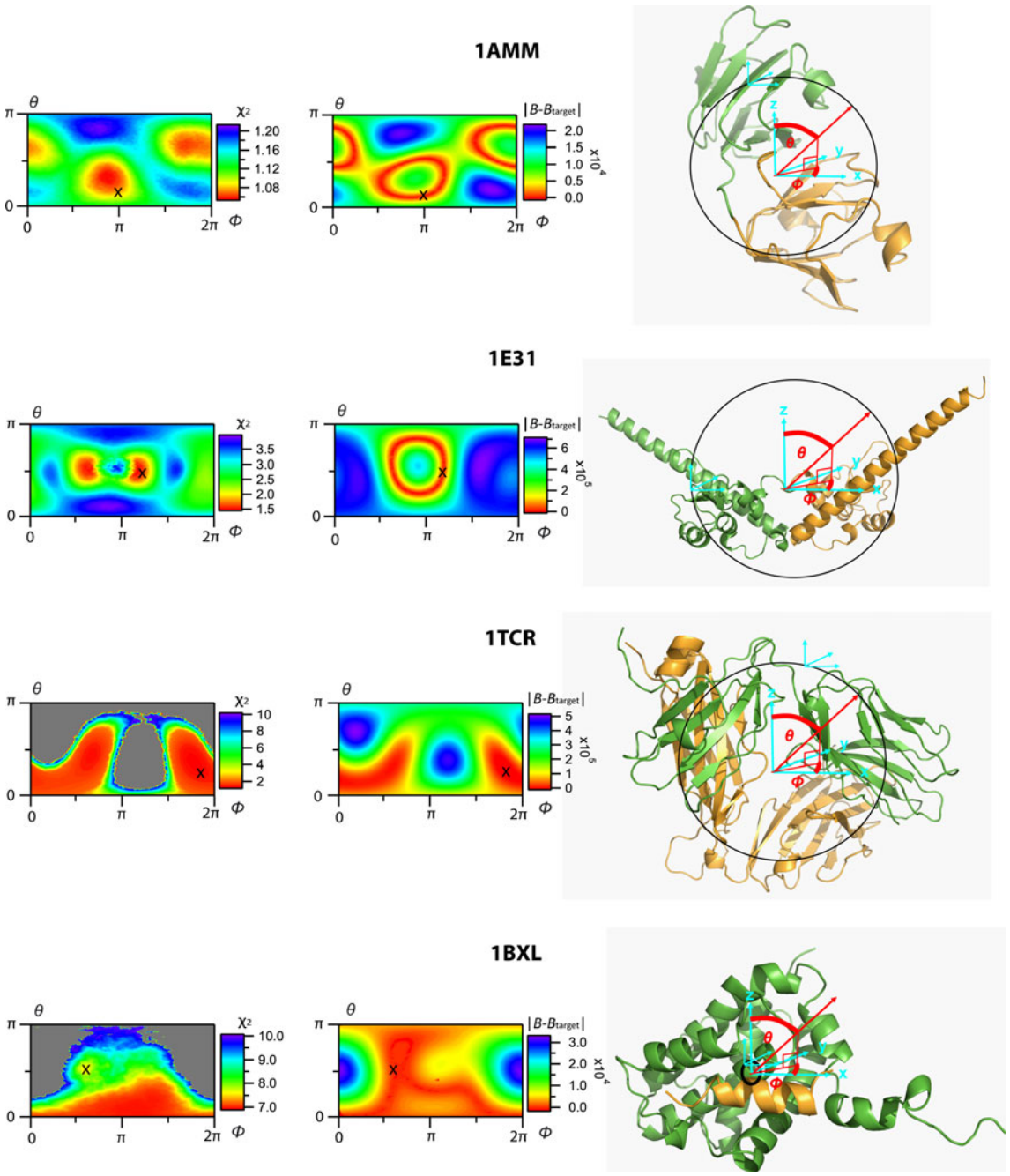

Fig. 2 χ^2 values and $|B - B_{target}|$ for different PDB structures (Table 1) as a function of the subunit 1 (*green*) center in the θ/Φ-space. *Red regions* indicate lowest χ^2 (*left column*) or $|B - B_{target}|$ values (*middle column*). The regions in *violet* indicate poorest χ^2 fits or largest $|B - B_{target}|$ values, respectively. *Grey areas* indicate $\chi^2 > 10$. A *black cross* marks the coordinates of the target structure. The *right column* shows the target structures and the θ/Φ reference frames. The *circles*/*ellipsoids* indicate, symbolically, the subunit 1 mobility on a spherical surface around the complex center of mass

curve confines the possible subunit positions to specific regions on a sphere (the θ/Φ-space) with a diameter (= the inter-subunit distance) that is determined by the radius of gyration R_g. For all 10,000 conformations sampled, the B-values were calculated from the atomic coordinates by use of Eq. 2. The target values B_{target} were calculated from the initial (target) structures and the difference $|B - B_{target}|$ evaluated in each case and plotted as a function of θ and Φ. The advantages of this approach compared with the least χ^2-scoring procedure are twofold:

1. it is sufficient to use scattering data limited to a Q-range corresponding to 2–3 times the Guinier range (Gabel et al. 2006); and

2. the difference $|B - B_{target}|$ can be predicted analytically for the entire θ/Φ-space from the target structure and needs not be calculated for each individual model (Gabel et al. 2006). Its full mathematical expression is provided in the "Appendix".

Results

Systems composed of two simple geometric bodies

Several bi-ellipsoidal systems were studied, composed either of two symmetric ellipsoids with axial ratios 50 Å:25 Å:25 Å (Fig. 1) or of a 50 Å:25 Å:25 Å ellipsoid and a sphere with radius 25 Å (Suppl. Fig. 2) in several relative positions and orientations. The results of the χ^2-fits with the SAXS curves in the angular range 0…0.1 $Å^{-1}$ with low noise are shown as a function of the first subunit center position in the θ/Φ-space in Fig. 1 and Suppl. Fig. 2. The position of subunit one (in green) in the target structure complex is marked by an "*X*" in each case. In addition, the value of $|B - B_{target}|$ is shown in θ/Φ-space. Additional conditions treating different angular ranges and SAXS noise are included in the supplementary material.

The "rod-like" conformation (Fig. 1, first row) displays the most restrictive conditions on the translational degrees of freedom for the subunit positions within the complex: there exist only two conformations that are equivalent in terms of χ^2-fit against the SAXS target curve. Both are related to the symmetry of the system and correspond to a permutation of the subunits. Small translational motion away from the two solutions result in significant deterioration of the χ^2-values. The "rod-like" conformation is also the most accurate and efficient in terms of stability against variations in the noise-level of the SAXS data and the *Q*-ranges investigated: in all cases the two symmetric solutions were uniquely and easily identified as the lowest χ^2-solutions confined to very restricted regions in the θ/Φ-plane even at high noise (Suppl. Fig. 4). The rod-like bi-ellipsoid case could be described in an equivalent manner by using the $|B - B_{target}|$ approach without losing any accuracy in terms of restrictions in the θ/Φ-space (Fig. 1).

Both the "L-shaped" and the "T-shaped" complexes yielded four structural solutions that were equivalent to the respective target structure as a consequence of the symmetry in the systems (Fig. 1, second and third row). The L-shaped conformation was less restrictive than the T-shaped in respect of variations of subunit positions; smaller translational shifts were tolerated and yielded similar χ^2-fits against the target SAXS curve. The L-shaped conformation required a low noise-level and a larger *Q*-range ($\geq$0.2 $Å^{-1}$) in order to constrain the refined domain positions to small regions in the θ/Φ-space. When larger errors were present and a small *Q*-range was used, solutions with low χ^2-values were found widely spread over the θ/Φ-space (Suppl. Fig. 4). For the L-shaped complex the analysis in terms of the *B*-potential (Fig. 1) could be considered as equivalent to the fits against the scattering data in restricted *Q*-ranges (<0.2 $Å^{-1}$) at larger noise levels: it provided a band in θ/Φ-space of equivalent solutions that overlapped with the regions of lowest χ^2. At low noise levels and larger *Q*-values, the χ^2-analysis was more discriminative. In the case of the T-shaped complex, the limiting factor of the accuracy and efficiency of the structural refinement was the noise level of the SAXS data: the high noise data enlarged the θ/Φ-space of good solutions significantly, particularly when a small angular range was considered (Suppl. Fig. 4). In the T-shaped case, the *B*-potential analysis yielded solutions in a restricted conformational space (band) but could not neatly detect the fourfold degeneracy of the possible subunit positions.

The "side-by-side" conformation was the least determined one in terms of restrictions in the θ/Φ-space. An infinite number of structures existed yielding equivalent fits against the overall SAXS target curve. Again, this family of structures could be understood by symmetry arguments and corresponded to the degree of freedom of shifting one subunit around the other in a specific "orbit" (Fig. 1, fourth row). The side-by-side conformation was also the most "vulnerable" to increased experimental errors: in almost none of the experimental *Q*-ranges studied were the structural solutions confined to small regions in the θ/Φ-space (Suppl. Fig. 4). False positives with lower χ^2-values than the target structure were found in several regions in the θ/Φ-space at high noise levels suggesting artifacts in the scoring process, possibly because of an effect of the hydration shell optimized by CRYSOL. The analysis in terms of the *B*-potential yielded equivalent structural solutions as the least χ^2-fit procedure in the "side-by-side" case and no loss of structural information was observed.

A more general system with two distinct three-axial ellipsoids (axial ratios 40 Å:30 Å:20 Å and 50 Å:20 Å:10 Å) at non-specific mutual arrangement was also studied (Suppl. Fig. 6). The similarities and conclusions found between the χ^2 and the *B*-analyses in the specific bi-ellipsoidal systems were equally reproduced in this more general case.

Two-body protein complexes from the PDB

The four different protein complexes studied (PDB entries 1AMM, 1BXL, 1E31, and 1TCR) covered a wide range of relative subunit sizes and shapes (Table 1). The χ^2-fits against the SAXS curves calculated from the target structures and the $|B - B_{target}|$ term are juxtaposed for each

system in Fig. 2. The position of subunit one from the target structure is marked with an "*X*" in each θ/Φ-space diagram. In general, the target structures in all cases were located in the restricted, one-dimensional subspaces of minimum $|B - B_{\text{target}}|$ values (red stripes in Fig. 2, middle column) as predicted by theory (Gabel et al. 2006). In each case, the $|B - B_{\text{target}}|$ plot therefore enabled graphical visualization of the possible subunit positions that were in good agreement with the SAXS data up to Q^4 (see Eq. 1).

When the structures were scored against the full SAXS curves, the target structures were in general found amongst those of lowest χ^2-values (see illustrative examples in Suppl. Figs. 1a and 1b). However, some exceptions were observed: in the case of 1AMM, the target structure ($\theta/\Phi = 0.42/3.12$) had a χ^2-value of 1.063 with the scattering data. Slightly better solutions were spread near the region $\theta/\Phi = 1.1/2.8$ with χ^2 as low as 1.056 (Fig. 2, top). The existence of slightly better structures than the target structure can probably be explained by the fact that CRYSOL uses a 3 Å thick hydration layer of variable (scattering length) density and small variations in the solvent-excluded volume of amino acid residues to optimize the fit of a given model against a SAXS curve (Svergun et al. 1995). It should be noted, though, that many of these slightly better models could be discarded because of a finite linker length (Suppl. Fig. 5). As a consequence of the spherical shape of the two subunits, the χ^2 analysis of the full SAXS curve was not able to distinguish their possible positions very accurately, particularly when high experimental noise was present (Suppl. Fig. 1a, top). Importantly, the χ^2-fit and the $|B - B_{\text{target}}|$ approach yielded similar (although not identical) restraints in terms of restrictions in the overall θ/Φ-space (Fig. 2, top).

In the 1E31 and 1TCR systems, both approaches (χ^2-fit and $|B - B_{\text{target}}|$ analysis) performed equally well (Fig. 2). The 1E31 system required low noise levels in the SAXS data in order to retrieve the target structure accurately with a χ^2-scoring (Suppl. Fig. 3) while the best solutions in the $|B - B_{\text{target}}|$ analysis comprised the target structure more faithfully. In the 1TCR system, both analyses yielded comparable results (Fig. 2). The χ^2-scoring became more efficient when larger Q-ranges were taken into account ($Q > 0.2$ Å^{-1}) (Suppl. Fig. 3).

The 1BXL structure was particularly interesting: the χ^2-fit and $|B - B_{\text{target}}|$ analysis yielded similar results, however, the lowest χ^2-values were found in a region of the conformational space far away from the target structure (Fig. 2, fourth row, left column). A potential reason for these "false positives" was possibly the optimization used in CRYSOL of the hydration shell which was contributing particularly strongly in this case because of the limited size (1.7 kDa) of the small peptide partner in the complex. This case illustrates that in some cases SAXS data are not sufficient to identify the target structure as the best one, even when the respective subunit orientations are known (see also Suppl. Fig. 1b, bottom). In these cases, additional restraints such as a mapping of the binding interface are necessary (Sattler et al. 1997).

The optional activation of steric restraints reduces the possible structural models significantly. Their effect is illustrated by two examples (Suppl. Fig. 5):

1. the finite length of a connecting linker of 7 amino acid residues reduces the conformational space of 1AMM structural models by more than 50%; and
2. the interdiction of domain interpenetration discards some of the potential models in the case of 1E31, notably many with already poor χ^2-values.

The effect of subunit deformation on binding was investigated by an example classified as "difficult" (RMSD 3.41) in the Benchmark library (Hwang et al. 2008): the 1B39:1FPZ complex forming 1FQ1 (Suppl. Fig. 8). The *B*-plots comparing the two free subunits as rigid bodies and the refined complex revealed that the procedure is stable against minor conformational changes of the subunits during complex formation.

The effect of uncertainties in the subunit orientation of some degrees (as expected from experimental RDC data) was probed in the 1E31 system by imposing a generous 10° flexibility around the three Cartesian axes of the first subunit. The back-calculated *B*-plots show that the procedure is stable against orientational uncertainties of that order (Suppl. Fig. 9).

Errors in the inter-subunit distance have been probed in the case of 1E31 by imposing artificially a wrong R_g-value of some percent (Suppl. Fig. 10). The results show that the procedure is dependent on an accurate (±2–3%) determination of the radius of gyration.

Discussion

In the early days of rigid-body modeling, educated guesses about possible respective orientations and positions were studied by using simple geometric bodies to build protein complexes (Glatter and Kratky 1982; Kratky 1972). More recent approaches used high-resolution models of the subunits in combination with grid searches to optimize domain positions or binding angles (Konarev et al. 2001; Mattinen et al. 2002; Nöllmann et al. 2004). The difficulties of uniquely refining two-body systems beyond their inter-subunit distance were discussed early in the framework of the triangulation method (Engelman and Moore 1972; Hoppe 1973; Hoppe et al. 1975; Kratky and Worthmann 1946; Moore and Engelman 1977) or in terms of variations of the pair distance-distribution function (Glatter 1979).

The approach presented here can be regarded as a *generalization of the triangulation method* going beyond the determination of the inter-subunit distance by proposing possible positions for the subunits at a given, fixed, inter-subunit distance. This improvement with respect to the above-mentioned techniques has become possible because the relative subunit orientations within a complex can nowadays often be determined by NMR residual dipolar couplings (RDCs), a technique that has been used with increasing success over the last 15 years (Bax 2003; Tjandra and Bax 1997).

Comparison of the generalized triangulation method and the least χ^2-scoring procedure

The overall comparison of the two approaches demonstrates that in many cases studied the $|B - B_{\text{target}}|$ analysis is a valid and quick alternative to the least χ^2-fit (using the full SAXS curve) showing the spread of the best structures in the conformational space (Figs. 1, 2; Suppl. Fig. 2). These intriguing findings show that the quality (*accuracy*, *uniqueness, and stability*) of rigid-body modeling can be judged and visualized in many cases in a very satisfactory way by $|B - B_{\text{target}}|$ analysis.

While the *B*-analysis is not very efficient in pinpointing specific subunit positions (it yields restrictions to one-dimensional lines in the θ/Φ-plane) it has two major advantages:

1. it is limited to an analysis of the lowest angles (in practice 2–3 times the Guinier range; Gabel et al. 2006); and
2. it is faster and more elegant than the χ^2-fit because the expressions of $|B - B_{\text{target}}|$ can be directly calculated analytically from a single model (e.g. the putative target structure) for all θ/Φ-coordinates (see "Appendix") whereas the χ^2-analysis requires that all models be explicitly generated and scored.

Even though the χ^2-analysis is often more successful in identifying and pinpointing subunit positions (in particular if experimental noise is low and large *Q*-ranges are available) it has its limits (Suppl. Figs. 1a and 1b):

1. As some of the PDB examples showed, a fit against the overall SAXS curve with adjustable hydration shell and solvent-excluded volumes of amino acid residues may lead to slightly false positives. In other words the target structure may not be identified as the "best" structure and represents only a side minimum of χ^2-values. The effect is most pronounced in systems with low anisometry (1AMM) or when one of the two partners is very small (1BXL).
2. As in the case of the $|B - B_{\text{target}}|$ analysis, the χ^2-analysis is limited by the presence of symmetry (Fig. 1) and several equivalent solutions may exist that cannot be distinguished.

Given these results the author believes that the $|B - B_{\text{target}}|$ analysis is a valid and quick alternative to the χ^2-analysis that can predict, judge, and visualize the expected refinement efficiency for a given system. However, an important requirement and limiting factor for de novo structure determination is the accurate determination (±2–3%) of the radius of gyration of the system under investigation.

The effect of subunit shape, distance, and orientation on the refinement capacity in rigid-body modeling

As illustrated by the examples (in particular the bi-ellipsoidal systems) the mere presence of a rather high degree of anisometry in a single (or both) subunit(s) in a complex is a necessary but not sufficient criterion to evaluate the refinement efficiency of the system by using SAS data. The quality of the refined model depends, in general, significantly on the relative position and orientation of both subunits. This was particularly evident in the case of rotational ellipsoidal bodies where the overall "rod-like", "T-shaped", and "L-shaped" target conformations yielded rather specific and restrictive restraints on the possible conformational degrees of freedom of the subunits while the "side-by-side" target conformation allowed for an infinite number of structural solutions because of the symmetry of the system.

While a high $|B - B_{\text{target}}|/B_{\text{target}}$ value (Table 1) is a necessary requirement for successful refinement (i.e. the scattering curve varies significantly as a function of subunit positions) it is not sufficient to guarantee its uniqueness (e.g. "side-by-side" case). The author therefore proposes use of the $|B - B_{\text{target}}|$ plots (Figs. 1, 2; Suppl. Fig. 2) as a quality criterion. Preparation and inspection of these plots can be performed quickly and easily by the experimenter to provide a quantitative criterion to evaluate the success and reliability of structural refinement *before* sophisticated model-building procedures are initiated. The plots can even be useful *before* SAS and NMR experiments for complexes of known individual subunit structures in order to anticipate and judge the potential outcome and efficiency of rigid-body modeling by probing putative subunit arrangements: in this case the user would sample several relative orientations at reasonable inter-subunit distances to obtain an idea of the uniqueness of the planned rigid-body refinement. Given the complexity and costs in terms of biochemistry work, SAS and NMR experiments as well as sophisticated data modeling, this approach is very attractive as a preliminary study. A third possibility to use the procedure is to check a structural model a posteriori to sophisticated

refinement procedures in order to judge the confidence that can be put into its stability, in other words to check if alternative solutions with similar goodness of fit against the SAS curve may exist that have been overlooked and to visualize the ensemble of possible solutions graphically.

As illustrated in Suppl. Fig. 7, the procedure is efficient in visualizing the ensemble of good models for sufficiently anisometric subunits even if the exact radius of gyration of the complex is not matched in the model calculations. Small deformations of the subunits upon complex formation (Suppl. Fig. 8) do not affect the results significantly. It is obvious, however, that the procedure (as any rigid-body modeling procedure) will lead to erroneous interpretations if the scale of these changes is too important. Small orientational uncertainties of some degrees of the subunits are equally tolerated (Suppl. Fig. 9). It should be noted, however, that the procedure yields very different results if subunits are flipped by 180° (degeneracy of domain orientations from experimental RDCs from a single alignment medium). Therefore, if it is important in a specific case to learn about the variability of structural solutions related to a specific model, it is necessary to lift the orientational degeneracy that is inherent in a single RDC data set by combining data from multiple alignment media (Bax 2003).

The most critical point of the procedure (if experimental SAS data are interpreted) is its sensitivity against errors in the radius of gyration (leading to an error in the *A*-term or inter-subunit distance). Although an imprecise inter-domain distance does not affect the distribution of good structural solutions in the θ/ϕ-space if the *B*-value is adapted to it (Suppl. Fig. 7), a simultaneous error of both the *A* and *B*-terms (of opposite directions) can yield erroneous interpretations and does not enable the target model to be retrieved. For the procedure to work accurately, it is therefore important that the radius of gyration is known with an accuracy of approximately 2–3% (e.g. 40 ± 1 Å) and the *B*-term with an accuracy of approximately 5% (Suppl. Fig. 10). This tolerance may be increased by a few percent (similar to the observations in Suppl. Fig. 7) if the errors of the *A* and *B*-terms (extracted with an alternating polynomial fit (Eq. 1) from experimental data) are in the same direction and compensate.

Conclusions

Rigid-body modeling of bio-macromolecular complexes composed of two or more subunits of known high-resolution structure in combination with non-crystallographic techniques (electron microscopy, small-angle scattering, NMR…) is becoming increasingly important to a growing community of structural biologists. A crucial point of these multi-disciplinary methods is the accuracy, uniqueness, and stability of the models refined. The procedure presented in this work enables quick and easy evaluation of the efficiency of bio-macromolecular rigid-body modeling using SAS and NMR restraints by visualizing the conformational space of possible models in agreement with the data. It is stable against small errors in subunit orientation and deformation upon complex formation but is sensitive to accurate determination of the inter-subunit distance. It can be used either before the actual structural refinement, to anticipate the quality of the expected models as a function of subunit geometry and sizes, or a posteriori to a structural refinement process, to visualize the conformational space of alternative structural solutions.

Appendix: Mathematical expression of $|B - B_{target}|$

The following mathematical expression has been adopted from Gabel et al. 2006. It is calculated from a specific model composed of two bodies K_1 and K_2 with inter-subunit distance $R_1 + R_2$. The sums of moments of the Cartesian coordinates x_i, y_i,, and y_i run within the respective domains separately and are calculated with respect to their respective centers of scattering length density. These moments need to be calculated only once for the procedure (e.g. from the putative target model). The trigonometric expressions enable prediction of the *B*-value from a specific model to the entire θ/Φ-space.

$$
\begin{aligned}
&B(\theta,\phi)\\
&= 4(R_1+R_2)^2\left[\begin{array}{l}\sin^2\theta(a_B\cos^2\phi + b_B\sin^2\phi + c_B\sin\phi\cos\phi)\\ \quad + d_B\cos^2\theta + \sin\theta\cos\theta(e_B\cos\phi + f_B\sin\phi)\end{array}\right]\\
&\quad + 4(R_1+R_2)[\sin\theta(g_B\cos\phi + h_B\sin\phi) + i_B\cos\theta],\\
&a_B = \sum_{i\in k_1, j\in k_2}\left(x_i^2 + x_j^2\right),\quad b_B = \sum_{i\in k_1, j\in k_2}\left(y_i^2 + y_j^2\right),\\
&\quad c_B = 2\sum_{i\in k_1, j\in k_2}(x_i y_i + x_j y_j),\\
&d_B = \sum_{i\in k_1, j\in k_2}\left(z_i^2 + z_j^2\right),\quad e_B = 2\sum_{i\in k_1, j\in k_2}(x_i z_i + x_j z_j),\\
&\quad f_B = 2\sum_{i\in k_1, j\in k_2}(y_i z_i + y_j z_j),\\
&g_B = \sum_{i\in k_1, j\in k_2}\left(x_i^3 + y_i^2 x_i + z_i^2 x_i - x_j^3 - y_j^2 x_j - z_j^2 x_j\right),\\
&h_B = \sum_{i\in k_1, j\in k_2}\left(x_i^2 y_i + y_i^3 + z_i^2 y_i - x_j^2 y_j - y_j^3 - z_j^2 y_j\right),\\
&i_B = \sum_{i\in k_1, j\in k_2}\left(x_i^2 z_i + y_i^2 z_i + z_i^3 - x_j^2 z_j - y_j^2 z_j - z_j^3\right),
\end{aligned}
$$

References

Back JW, de Jong L, Muijsers AO, de Koster CG (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. J Mol Biol 331:303–313

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39:5355–5365

Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

Bernadó P, Blackledge M (2010) Structural biology: proteins in dynamic equilibrium. Nature 468:1046–1048

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog Nucl Magn Reson Spectrosc 46:23–46

Capel MS, Engelman DM, Freeborn BR, Kjeldgaard M, Langer JA, Ramakrishnan V, Schindler DG, Schneider DK, Schoenborn BP, Sillers IY, Yabuki S, Moore BP (1987) A complete mapping of the proteins in the small ribosomal subunit of Escherischia coli. Science 238:1403–1406

Chantalat L, Skoufias DA, Kleman JP, Jung B, Dideberg O, Margolis RL (2000) Crystal structure of human survivin reveals a bow tie-shaped dimer with two unusual alpha-helical extensions. Mol Cell 6:183–189

Debye P (1915) Zerstreuung von Roentgenstrahlen. Ann Physik 28:809–823

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737

Engelman DM, Moore PB (1972) A new method for the determination of biological quaternary structure by neutron scattering. Proc Natl Acad Sci USA 69:1997–1999

Gabel F, Simon B, Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. Eur Biophys J 35:313–327

Gabel F, Simon B, Nilges M, Petoukhov M, Svergun D, Sattler M (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. J Biomol NMR 41:199–208

Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA (1996) An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science 274:209–219

Glatter O (1979) The interpretation of real-space information from small-angle scattering experiments. J Appl Crystallogr 12:166–175

Glatter O, Kratky O (1982) Small angle X-ray scattering. Academic Press, London

Goldstein H (1977) Classical mechanics. Addison-Wesley, Reading

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127:16621–16628

Heller WT (2010) Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. Acta Crystallogr Sect D Biol Crystallogr 66: 1213–1217

Hoppe W (1973) The label triangulation method and mixed isomorphous replacement principle. J Mol Biol 78:581–585

Hoppe W, May R, Stoeckel P, Lorenz S, Erdmann VA, Wittmann HG, Crespi HL, Katz JJ, Ibel K (1975) Neutron scattering measurements with the label triangulation method on the 50S subunit of *E. Coli* ribosomes. In: Schoenborn BP (ed) Neutron scattering for the analysis of biological structures, vol 27. Brookhaven National Laboratory, Brookhaven, pp IV38–IV48

Hwang H, Pierce B, Mintseris J, Janin J, Weng Z (2008) Protein-protein docking benchmark version 3.0. Proteins 73:705–709

Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. Protein Sci 19:642–657

Jacrot B (1976) The study of biological structures by neurtron scattering from solution. Rep Prog Phys 39:911–953

Konarev PV, Petoukhov MV, Svergun DI (2001) MASSHA—a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. J Appl Crystallogr 34:527–532

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55

Kratky O (1972) Recent advances and applications of diffuse X-ray small-angle scattering on biopolymers in dilute solutions. Q Rev Biophys 5:481–537

Kratky O, Worthmann W (1946) Ueber die Bestimmabrkeit der Konfiguration geloester organischer Molekule durch interferometrische Vermessung von Roentgenstrahlen. Monatsh Chem 76:263–281

Kumaraswamy VS, Lindley PF, Slingsby C, Glover ID (1996) An eye lens protein-water structure: 1.2 A resolution structure of gammaB-crystallin at 150 K. Acta Crystallogr Sect D 52: 611–622

Lipfert J, Doniach S (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu Rev Biophys Biomol Struct 36:307–327

Macheboeuf P, Piuzzi M, Finet S, Bontems F, Pérez J, Dessen A, Vachette P (2011) Solution X-ray scattering study of a full-length class A penicillin-binding protein. Biochem Biophys Res Commun 405:107–111

Madl T, Gabel F, Sattler M (2011) NMR and small-angle scattering-based structural analysis of protein complexes in solution. J Struct Biol 173:472–482

Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand J-Y, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data. Eur Biophys J 37:95–104

Marino M, Zou P, Svergun D, Garcia P, Edlich C, Simon B, Wilmanns M, Muhle-Goll C, Mayans O (2006) The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from Tintin. Structure 14:1437–1447

Mattinen M-L, Paakkonen K, Ikonen T, Craven J, Drakenberg T, Serimaa R, Waltho J, Annila A (2002) Quaternary structure built from subunits combining NMR and small-angle X-Ray scattering data. Biophys J 83:1177–1183

Moore PB, Engelman DM (1977) Model calculations of protein pair interference functions. J Mol Biol 112:228–234

Nöllmann M, He J, Byron O, Stark WM (2004) Solution structure of the Tn3 resolvase-crossover site synaptic complex. Mol Cell 16:127–137

Petoukhov MV, Svergun DI (2007) Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr Opin Struct Biol 17:562–571

Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P, Fernández-Recio J (2010) Structural characterization of protein–protein complexes by integrating computational docking with small-angle scattering data. J Mol Biol 403:217–230

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285

Sattler M, Liang H, Nettesheim D, Meadows RP, Harlan JE, Eberstadt M, Yoon HS, Shuker SB, Chang BS, Minn AJ, Thompson CB, Fesik SW (1997) Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. Science 275: 983–986

Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, Clore GM (2010) Solution structure of the 128 kDa enzyme I

dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc 132:13026–13045

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76:2879–2886

Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28:768–773

Takayama Y, Schwieters CD, Grishaev A, Ghirlando R, Clore GM (2011) Combined use of residual dipolar couplings and solution X-ray scattering to rapidly probe rigid-body conformational transitions in a non-phosphorylatable active-site mutant of the 128 kDa enzyme I dimer. J Am Chem Soc 133:424–427

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114

Wang J, Zuo X, Yu P, Byeon IJ, Jung J, Wang X, Dyba M, Seifert S, Schwieters CD, Qin J, Gronenborn AM, Wang YX (2009a) Determination of multicomponent protein structures in solution using global orientation and shape restraints. J Am Chem Soc 131:10507–10515

Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX (2009b) A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. J Mol Biol 393: 717–734

Yuzawa S, Ogura K, Horiuchi M, Suzuki NN, Fujioka Y, Kataoka M, Sumimoto H, Inagaki F (2004) Solution structure of the tandem Src homology 3 domains of pH47phox in an autoinhibited form. J Biol Chem 28:29752–29760

Zuiderweg ERP (2002) Mapping protein–protein interactions in solution by NMR spectroscopy. Biochemistry 41:1–7